Analysis of methylomic data in .fastq files using WG-Blimp

Ramanuj Sarkar and Patrick Allard
Molecular Biology Institute, UCLA

Abstract:

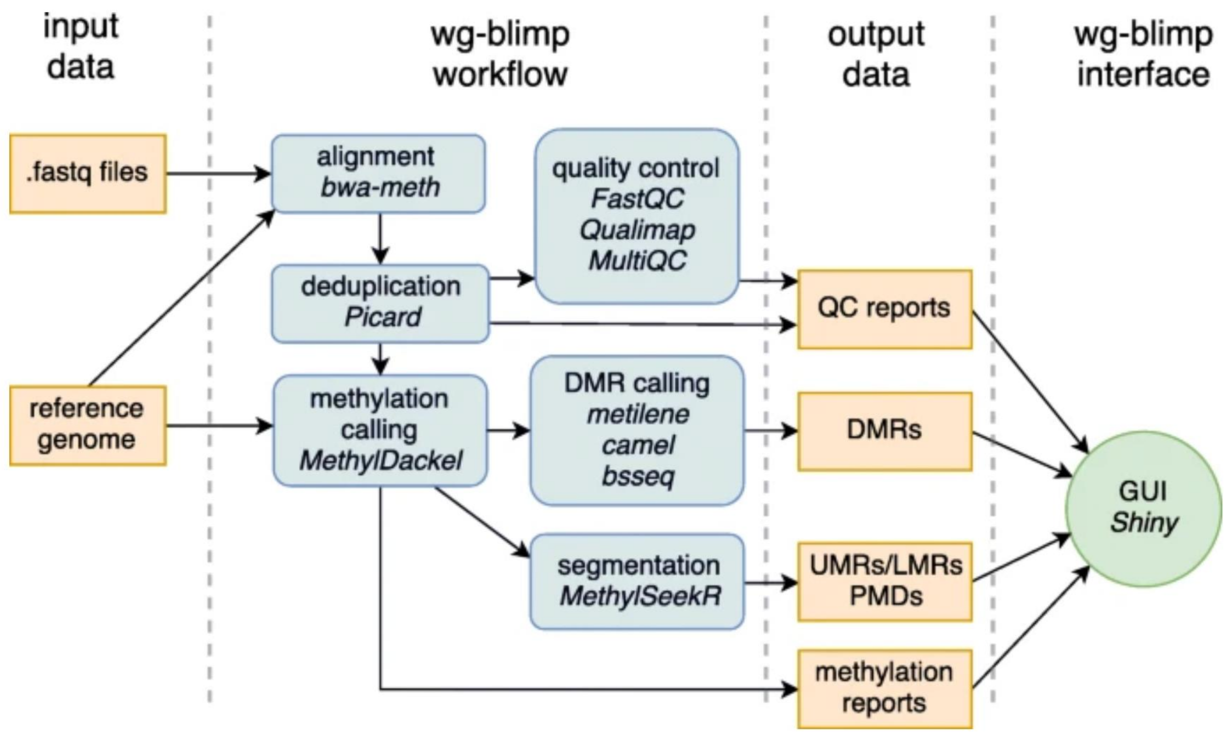
Bisphenol A (BPA) has been associated with changes to the frequency of gene expression in mouse epiblast-like cells (EpiLCs) without changing the effect when compared to control EpiLCs. This project analyzed pre-collected methylomic data from the two groups of EpiLCs and identified regions with different methylation between them to understand if the methylation corresponded with the changed frequencies. We attempted to use the end-to-end analysis pipeline wg-blimp to create methylation reports which would convey whether these were correlated. The results were inconclusive.

Introduction:

The cells created by genes may act differently, which includes changes to the actual DNA (genetic) or changes to which parts of the DNA are read and how often (epigenetic). One epigenetic change to cells is when they are methylated, in which case a methyl group (CH3) is attached to the base.

Bisphenol A is a molecule used in polycarbonate plastics. When mouse epiblast-like cells (EpiLCs), which came from mice and mimicked the state of early embryonic cells, were exposed to Bisphenol A, the gene expression within them was similar to EpiLCs which had not been exposed[1]. However, the control and BPA-exposed cells had different levels of expression of certain genes.

A technique called whole-genome bisulfite sequencing allows people to better understand which parts of the DNA are methylated by turning all methylated cytosine bases into uracil, which acts differently in a way which can be observed in .fastq files. The end-to-end pipeline wg-blimp creates a GUI with which one can view data from other programs in its workflow, taking .fastq files and the reference genome as input, then outputting data such as methylation reports which can be used to determine the methylation patterns on the DNA in the .fastq files.[2]



A visual representation of the wg-blimp workflow and interface[2]

It is possible to install the programs which make up the wg-blimp workflow, such as MethylDackel and metilene, without installing wg-blimp, such as by using the package management tool bioconda. It is recommended that, if wg-blimp is downloaded using bioconda, the download should occur in a fresh environment.[2]

Methods:

The object of this B.I.G. Summer 2022 Program project was downloading wg-blimp and using it to analyze the .fastq files. All of the installations of the wg-blimp in this experiment were attempted on a Lenovo desktop running Windows 11, via Windows Subsystem for Linux Preview (WSL) using the Linux Ubuntu OS.

The first attempted installation used bioconda in an environment where metilene was also present. This was due to a pre-existing While downloading wg-blimp, the computer crashed and the attempt was therefore unsuccessful.

The second attempted installation was an installation from source. The programs metilene, bwa-meth, Picard, and MethylDackel were successfully downloaded separately in that order. While downloading MethylDackel, the computer crashed twice before installing successfully the third time. Downloads of fastqc and MethylSeekR were attempted, but neither of these downloads were successful. It is possible this was due to the computer crashing, even though a successful install was indicated on the WSL command line. Therefore, the second installation attempt was unsuccessful.

The third attempted installation occurred after uninstalling and reinstalling WSL and Ubuntu to create a fresh environment which only the package manager used to download wg-blimp was present. The download occurred overnight, but appeared to have resulted in a crash. This is because the computer was set to never sleep, and the screen was set to never turn off, but the computer indicated that it restarted. Therefore, the third installation attempt was unsuccessful

Results:

The results of the experiment were inconclusive, as wg-blimp was not able to be downloaded in any of the environments previously mentioned in the experiment. By the time the final environment was attempted, the allotted time span of the BIG Summer 2022 Project was nearly over.

Works Cited:

1 Ooi SKT, Jiang H, Kang Y, Allard P. 2021. Examining the Developmental Trajectory of an in Vitro Model of Mouse Primordial Germ Cells following Exposure to Environmentally Relevant Bisphenol A Levels. Environmental Health Perspectives. 129(9):097013. doi:10.1289/EHP8196. [accessed 2021 Nov 17]. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8480152/.

2 Wöste M, Leitão E, Laurentino S, Horsthemke B, Rahmann S, Schröder C. 2020. wg-blimp: an end-to-end analysis pipeline for whole genome bisulfite sequencing data. BMC Bioinformatics. 21(1). doi:10.1186/s12859-020-3470-5.